IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 11, 2018, accepted February 5, 2018, date of publication February 27, 2018, date of current version April 4, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2810216

Energy-Efficient Resource Allocation for Heterogeneous Wireless Network With Multi-Homed User Equipments

RUNZI LIU, (Member, IEEE), MIN SHENG, (Senior Member, IEEE), AND WEIHUA WU
State Key Laboratory of ISN, School of Telecommunications Engineering, Xidian University, Xi'an 710071, China

Corresponding author: Weihua Wu (xdweihuawu@hotmail.com)

This work was supported in part by the Natural Science Foundation of China under Grant 61701365, Grant 91638202, and Grant 61725103, in part by the China Postdoctoral Science Foundation under Grant 2017M623121, in part by the Fundamental Research Funds for the Central Universities under Grant JB170105, and in part by the Key Research and Development Program—The Key Industry Innovation Chain of Shaanxi under Grant 2017ZDCXL-GY-04-04.

ABSTRACT In this paper, we investigate the energy-efficient resource allocation problem for heterogeneous wireless network with multihomed user equipments. First, the energy-efficient resource allocation is formulated as an energy efficiency (EE) maximization problem, which is a mixed-integer nonlinear optimization (MINO) problem. We first introduce a continuity relaxation and Lagrange dual method to solve the MINO problem, which has relatively high computational complexity. For reducing the computational complexity of the resource allocation problem, we further propose a two-phase optimization method. Specifically, it includes the minimum rate guaranteed resource allocation and the energy-efficient resource allocation. Finally, we show that the two-phase optimization method achieves the suboptimal EE with significantly lower computational complexity compared with the optimal one. Simulation results show that the suboptimal algorithm achieves the EE performance higher than the conventional approach and comparable with the optimal one, but with much less complexity.

INDEX TERMS Wireless network resource allocation, energy efficiency, multi-homing.

I. INTRODUCTION

It is reported that there will be an astounding 1000-fold increase in media traffic generated by smartphones, tablets and machine-type communication devices for the wireless network in the last decades [1]. Although many lasted wireless technologies, such as WiMAX, Mobile-Fi, IEEE 802.11 wireless local area networks (WLAN) / Wi-Fi, and IEEE 802.15 wireless personal area networks (WPAN), have been proposed in recent years, supporting these media applications while maximizing the wireless network resource (e.g., power, bandwidth) utilization is a challenging task [2]. Specifically, these lasted wireless technologies are limited in radio coverage and mobility support for individual users [3]; while, the access point (AP) of traditional cellular network can well sustain the user mobility but its bandwidth is often inadequate to support the throughput-demanding video applications. To deliver high-quality media streaming service, it becomes vital to consider aggregating the bandwidth of heterogeneous wireless networks (HetNet). As a result, multi-homing service, where a user equipment (UE) maintains multiple simultaneous network paths between the media content server and the UE by employing different APs in the HetNet, is considered as a promising solution for offloading the massive media traffic [4].

Although multi-homing technology provides a lot of performances improvement, there are many technical challenges for realizing the multi-homing service in HetNet. First, due to the fact that each UE can maintain multiple simultaneous connections with different APs, the UEs QoS requirements deepen the coupling of the resource allocations of different APs, which leads to an across multiple APs resource allocation problem and hence increases the problem complexity. Second, most of the power energy of HetNet is consumed by the APs [5]. Moreover, 80% of these energy is dissipated as heat, and only around 5% to 20% of the input power is used for supporting the wireless traffic [6]. Therefore, in order to realize the environmental-friendly communication, energy efficiency (EE) must be a key performance metric for the resource allocation in multi-homing networks. However, the EE often leads to a non-convexity

EE function, so that the optimal energy-efficient resource allocation cannot be obtained by the conventional convex optimization algorithms. Third, deployed with the orthogonal frequency division multiple access (OFDMA) subcarrier, the allocation of these wireless resource is indicated by the binary allocation variable, which leads to a mixed-integer nonlinear programming [7]. According to the analysis above, the mentioned factors present difficult challenges to solve the energy-efficient resource allocation problem for the HetNet with multi-homing services.

In this paper, we first formulate the energy-efficient resource allocation as an EE maximization problem. We apply the nonlinear fractional programming to convert the EE maximization into a mixed integer nonlinear optimization (MINO) problem. Then, the optimal solution of MINO problem is derived by the continuity relaxation and Lagrange dual method, which has relatively high computational complexity. For reducing the computational complexity of the resource allocation problem, we also propose a two-phase optimization method, which includes the minimum rate guaranteed resource allocation problem, and the energy-efficient resource allocation problem. Finally, the two problems are solved orderly for obtaining the low-complexity resource allocation algorithm.

The main contributions of this paper are outlined as follows:

- We propose an across multiple APs resource allocation framework, with considering the UEs multi-homing capability.
- An iterative joint power and subcarrier allocation algorithm consisting of both outer and inner loop optimizations is proposed to achieve the globally optimal EE. Specifically, the algorithm is different from existing algorithms for OFDMA system in that the resource allocation is performed across multiple APs to maximize the system EE.
- To facilitate practical implementation, we also develop a low-complexity suboptimal joint power and subcarrier allocation algorithm. We prove that the proposed algorithm achieves the suboptimal EE with significantly lower computational complexity compared with the optimal one.

The rest of the paper is organized as follows: Section II presents the related works. The network model and problem formulation are given in Section III. In Section IV, we introduce the optimal energy-efficient resource allocation. Section V gives a low complexity suboptimal resource allocation algorithm. Section VI presents the simulation results to evaluate the proposed scheme. The conclusions are drawn in Section VII.

II. RELATED WORKS

Our work in this paper lies along the intersection of research contexts: 1) resource allocation for wireless networks, and 2) multi-homing media transmission.

There exists a large body of works conducted in resource allocation for wireless networks. For example, an airtime-based resource control technique was proposed in [8] for the virtualized wireless network, in which wireless network resources are allocated among competing virtual networks while keeping their programmability. The Bayesian-game based resource allocation schemes were developed in [9] to dictate the UE to request wireless resources based on the bidding strategies of other UE players. A three-phase search algorithm was developed in [10] to choose appropriate small cells and physical resources for UEs while minimizing the overall energy consumption and reducing the network interference. The resource allocation scheme in [11] develop an efficient distributed method to solve resource allocation problem. The problem of energy-aware resource management was formulated as a three-stage Stackelberg game in [12] and an iterative algorithm was proposed to obtain the Stackelberg equilibrium solution. In [8]–[12], all the resource allocation schemes have relatively high computational complexity, they cannot be realized within a short time window, which can be 10 ms for LTE system [13]. For reducing the computational complexity, the resource allocation scheme in [14] divided the original optimization problem into two subproblems, and the corresponding low-complexity algorithms were developed. The low-complexity heuristic algorithm was proposed in [15] to solve the resource allocation problem in OFDMA networks. Nevertheless, the low-complexity designs of [14] and [15] cannot be used in the HetNet with multi-homed UEs, where the coupling among multiple APs makes the resource allocation more complicated.

Multi-homing service has received growing attentions recent years. the piecewise linearization approach in [16] selectively dropped some packets under the battery energy limitation, and assigned the most valuable packets to different radio interfaces in order to minimize the video quality distortion. A traffic splitting scheme was developed in [17] for the cellular/WiFi heterogeneous networks to achieve various performance gains. Under the multi-homing video transmission scheme in [18], the UE can adapt its energy consumption to support at least the target video quality lower bound during the call. An optimal rate splitting strategy was developed in [19] to improve both spectral efficiency and energy efficiency by exploring and exploiting cooperation diversity. However, all of these works focused on the traffic scheduling among multiple APs, and the physical resource allocation were not explored.

III. SYSTEM MODEL

In this section, we first introduce the network model of HetNet including the definition of EE. Next, we discuss the problem formulation for energy-efficient resource allocation.

A. NETWORK MODEL

The considered HetNet owns N APs. An example system model of HetNet is shown in Fig. 1. Multiple radio interfaces

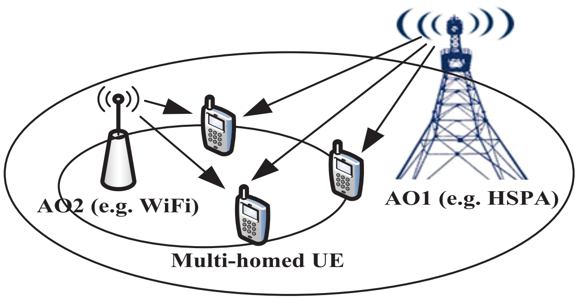


FIGURE 1. System model of HetNet with multi-homed UEs.

are deployed on the UE, so that they have multi-homing capable. With the help of these radio interfaces, the UE can establish communication with multiple APs simultaneously and employ them for media content transmission. We assume that, in HetNet, different APs operate in different bandwidths (e.g., 2.4 GHz for WiFi, 1.8-2.3 GHz for HSPA) [20]. Thus the inter-APs interference does not exist in the HetNet. Let $\mathcal{N} = \{1, 2, \cdots, N\}$ and $\mathcal{M} = \{1, 2, \cdots, M\}$ denote the set of APs and UEs in the system, respectively. Due to the practical limits in multi-homing capable, we consider that each UE m (AP n) can connect to the set of $\mathcal{N}_m$ APs ($\mathcal{M}_n$ UEs) from the available APs set $\mathcal{N}$ (UEs set $\mathcal{M}$), then $\mathcal{M} = \bigcup_{n=1}^{N} \mathcal{M}_n$, and $\mathcal{N} = \bigcup_{m=1}^{M} \mathcal{N}_m$.

The radio resources of APs include OFDM subcarrier and transmit power. The subcarrier set of AP n is denoted as $\mathcal{J}_n$, and its number is written as J_n. Binary variable $x_{n,m}^k$ indicates that subcarrier k in AP n is allocated to UE m. The channel gain between AP n and UE m on subcarrier k is $g_{n,m}^k$ and the transmit power is $p_{n,m}^k$. Then, the achievable data rate between AP n and UE m on subcarrier k can be expressed as

$$r_{n,m}^k = x_{n,m}^k \varepsilon_n B_n \log\left(1 + \Gamma \frac{|g_{n,m}^k|^2 p_{n,m}^k}{B_n N_0}\right), \quad (1)$$

where ε_n is the network efficiency depending on the decoder efficiency of an AP, B_n is the subcarrier spacing in AP n, Γ is the capacity gap from the Shannon channel capacity [7], and N_0 is the power spectral density of additive white Gaussian noise. Since subcarriers are allocated orthogonally among UEs in AP n, we have

$$\sum_{m \in \mathcal{M}_n} x_{n,m}^k \leq 1, x_{n,m}^k \in \{0, 1\}, \forall k \in \mathcal{J}_n, n \in \mathcal{N}. \quad (2)$$

Thus, the sum-rate for UE m, which is served by the set of $\mathcal{N}_m$ APs, is given as

$$R_m = \sum_{n \in \mathcal{N}_m} \sum_{k \in \mathcal{J}_n} x_{n,m}^k \varepsilon_n B_n \log\left(1 + \Gamma \frac{|g_{n,m}^k|^2 p_{n,m}^k}{B_n N_0}\right). \quad (3)$$

Let R_m^{req} denote the data requirement of UE $m \in \mathcal{M}$. In order to reconstruct the media content successfully, the rate-aggregation of the transmission rate on the multiple paths should be larger than R_{req}^m, so that the rate constraint is given as $R_m \geq R_{req}^m$, $m \in \mathcal{M}$.

B. ENERGY EFFICIENCY

Due to the fact that circuit components power consumption contributes to a large part of energy consumption in HetNet [21], the circuit power $p_{n,m}^c$ for establishing connection between AP n and UE m should be also considered in the total power consumption. Therefore, the total power consumption of AP n is expressed as

$$P_n = \sum_{m \in \mathcal{M}_n} \sum_{k \in \mathcal{J}_n} x_{n,m}^k p_{n,m}^k + \sum_{m \in \mathcal{M}_n} p_{n,m}^c, \quad \forall n \in \mathcal{N}. \quad (4)$$

In practice, the transmit power at the AP should not be unbounded, hence the transmit power must satisfies the following constraint

$$\sum_{m \in \mathcal{M}_n} \sum_{k \in \mathcal{J}_n} x_{n,m}^k p_{n,m}^k \leq P_n^{\max}, \forall n \in \mathcal{N}. \quad (5)$$

The EE is defined as the ratio of the total transmission rate to the corresponding total power consumption (unit: bits/watt):

$$\frac{C(\mathbf{x}, \mathbf{p})}{P(\mathbf{x}, \mathbf{p})} = \frac{\sum_{m \in \mathcal{M}} \sum_{n \in \mathcal{N}_m} \sum_{k \in \mathcal{J}_n} x_{n,m}^k \varepsilon_n B_n \log\left(1 + \Gamma \frac{|g_{n,m}^k|^2 p_{n,m}^k}{B_n N_0}\right)}{\sum_{n \in \mathcal{N}} \left(\sum_{m \in \mathcal{M}_n} \sum_{k \in \mathcal{J}_n} x_{n,m}^k p_{n,m}^k + \sum_{m \in \mathcal{M}_n} p_{n,m}^c\right)}, \quad (6)$$

where $\mathbf{x} = [\mathbf{x}_n]_{1 \times N}$, $\mathbf{x}_n = [x_{n,m}^k]_{|\mathcal{M}_n| \times J_n}$ [1] and $\mathbf{p} = [\mathbf{p}_n]_{1 \times N}$, $\mathbf{p}_n = [p_{n,m}^k]_{|\mathcal{M}_n| \times J_n}$ represent the feasible subcarrier and power allocation polices, respectively.

C. PROBLEM FORMULATION

With the related constraints, the energy-efficient resource allocation problem is formulated as

$$\max \quad \frac{C(\mathbf{x}, \mathbf{p})}{P(\mathbf{x}, \mathbf{p})} \quad (7)$$

subject to the following constraints:
C1 (Orthogonality constraint):

$$\sum_{m \in \mathcal{M}_n} x_{n,m}^k \leq 1, x_{n,m}^k \in \{0, 1\}, \forall k \in \mathcal{J}_n, n \in \mathcal{N}.$$

C2 (UEs QoS constraint):

$$\sum_{n \in \mathcal{N}_m} \sum_{k \in \mathcal{J}_n} x_{n,m}^k \varepsilon_n B_n \log\left(1 + \Gamma \frac{|g_{n,m}^k|^2 p_{n,m}^k}{B_n N_0}\right) \geq R_m^{req}, \quad \forall m \in \mathcal{M}.$$

C3 (Total power constraint):

$$\sum_{m \in \mathcal{M}_n} \sum_{k \in \mathcal{J}_n} x_{n,m}^k p_{n,m}^k \leq P_n^{\max}, \quad \forall n \in \mathcal{N}.$$

We observe that the objective in optimization (7) is a fraction function, which results a non-convex optimization

[1] $|\mathcal{S}|$ is the cardinality of set $\mathcal{S}$.

problem. Moreover, due to the binary variables $\mathbf{x}$, the problem falls a mixed integer nonlinear optimization (MINO) problem, which is NP-hard generally [22]. Finally, the UEs QoS constraint makes the resource allocation of different APs coupled with each other, which increases the problem complexity.

IV. OPTIMAL ENERGY-EFFICIENT RESOURCE ALLOCATION

For coping with the non-convex optimization problem, the non-linear fractional programming is utilized in this section for converting the non-convex objective function in problem (7) into a differential form. Then, an iteration algorithm is developed to solve this EE maximization problem.

We first define the EE performance of the HetNet as a non-negative variable $\eta = C(\mathbf{x}, \mathbf{p})/P(\mathbf{x}, \mathbf{p})$. The optimal EE is defined as $\eta^{opt} = \max\limits_{\mathrm{C1-C4}} \frac{C(\mathbf{x},\mathbf{p})}{P(\mathbf{x},\mathbf{p})}$. Then, it comes the following lemma.

Lemma 1: The optimal EE of the resource allocation problem can be achieved if and only if

$$\begin{aligned} \max_{C1-C3} \quad & C(\mathbf{x}, \mathbf{p}) - \eta^{opt} P(\mathbf{x}, \mathbf{p}) \\ & = C(\mathbf{x}^*, \mathbf{p}^*) - \eta^{opt} P(\mathbf{x}^*, \mathbf{p}^*) \\ & = 0, \end{aligned}$$

where $\mathbf{x}^*$ and $\mathbf{p}^*$ are the optimal solutions of optimization problem (7).

Proof: Similar proof can be found in [23]. ■

It is shown in **Lemma 1** that if the optimal value η^{opt} is given, optimization problem (7) can be converted into the following optimization problem.

$$\begin{aligned} &\max\ C(\mathbf{x}, \mathbf{p}) - \eta^{opt} P(\mathbf{x}, \mathbf{p}) \\ &\text{s.t. } \mathrm{C1 - C4}. \end{aligned} \tag{8}$$

In fact, we can not known η^{opt} in advance. Hence, an iterative procedure is proposed in **Algorithm 1** to update η. With the definition of $F(\eta) = \max\limits_{\mathrm{C1-C4}} C(\mathbf{x}, \mathbf{p}) - \eta P(\mathbf{x}, \mathbf{p})$, solving optimization problem (8) is converted into finding the root for nonlinear equation $F(\eta) = 0$.

Lemma 2: Based on an iteration algorithm, $F(\eta)$ can converge to zero with a linear convergence rate.

Proof: Similar proof is shown in [23]. ■

Actually, there are two nested loops executed in **Algorithm 1**. On the one hand, η^{i+1} is iterated with the obtained $\mathbf{x}^i$ and $\mathbf{p}^i$ at the Outer Loop; on the other hand, $\mathbf{x}^i$ and $\mathbf{p}^i$ are solved through problem (9) at the Inner Loop.

$$\begin{aligned} &\max\ C(\mathbf{x}, \mathbf{p}) - \eta P(\mathbf{x}, \mathbf{p}) \\ &\text{s.t. } \mathrm{C1 - C4}. \end{aligned} \tag{9}$$

Now, resource allocation problem (7) is converted into solving problem (9) at given η. However, we found that problem (9) is an MINO problem, which is difficult to solve in generally. In the following, we will employ the primal-dual decomposition method in [7] and [24]. Moreover, the optimal solution can be derived based on the special structure of our optimization problem.

Algorithm 1 EE-Based Resource Allocation

Initialization Set the parameters about maximum iteration number $I_{\max}$, convergence condition ϵ and iteration index $i = 1$;
Set $\eta^1 = 0$ and begin iteration (Outer Loop);
for $1 \le i \le I_{\max}$ **do**
  Solve resource allocation problem (9) with η^k (Inner Loop);
  obtain $\mathbf{x}^i$, $\mathbf{p}^i$, $C(\mathbf{x}, \mathbf{p})$ and $P(\mathbf{x}, \mathbf{p})$;
  if $|C(\mathbf{x}^i, \mathbf{p}^i) - \eta^k P(\mathbf{x}^i, \mathbf{p}^i)| \le \epsilon$ **then**
    Set $\{\mathbf{x}^*, \mathbf{p}^*\} = \{\mathbf{x}^i, \mathbf{p}^i\}$ and $\eta^{opt} = \eta^i$;
      break
  else
    Set $\eta^{i+1} = \frac{C(\mathbf{x}^i,\mathbf{p}^i)}{P(\mathbf{x}^i,\mathbf{p}^i)}$ and $i = i + 1$;
  end if
end for

The primal-dual decomposition method includes the following two steps:

1) CONTINUITY RELAXATION

Relaxing $x^k_{n,m}$ to the continuous interval [0, 1] and introducing a new variable $s^k_{n,m} = x^k_{n,m} p^k_{n,m}$, problem (9) can be rewritten as

$$\begin{aligned} \max \sum_{m\in\mathcal{M}} \sum_{n\in\mathcal{N}_m} \sum_{k\in\mathcal{J}_n} x^k_{n,m} \varepsilon_n B_n \log\left(1 + \Gamma \frac{|g^k_{n,m}|^2 s^k_{n,m}}{x^k_{n,m} B_n N_0}\right) \\ - \eta \sum_{n\in\mathcal{N}} \left(\sum_{m\in\mathcal{M}_n} \sum_{k\in\mathcal{J}_n} s^k_{n,m} + \sum_{m\in\mathcal{M}_n} p^c_{n,m} \right) \end{aligned} \tag{10}$$

subject to the following constraints:

$$\begin{aligned} &\mathrm{C1}^*: \sum_{m\in\mathcal{M}_n} x^k_{n,m} \le 1,\ \ 0 \le x^k_{n,m} \le 1, ,\ \forall k \in \mathcal{J}_n,\ n \in \mathcal{N}, \\ &\mathrm{C2}^*: \sum_{n\in\mathcal{N}_m} \sum_{k\in\mathcal{J}_n} x^k_{n,m} \varepsilon_n B_n \log\left(1 + \Gamma \frac{|g^k_{n,m}|^2 s^k_{n,m}}{x^k_{n,m} B_n N_0}\right) \ge R^{req}_m, \\ &\qquad\qquad \forall m \in \mathcal{M}, \\ &\mathrm{C3}^*: \sum_{m\in\mathcal{M}_n} \sum_{k\in\mathcal{J}_n} s^k_{n,m} \le P^{\max}_n,\ \forall n \in \mathcal{N}, \\ &\mathrm{C4}^*: s^k_{n,m} \ge 0,\ \forall n \in \mathcal{N},\ \forall k \in \mathcal{J}_n,\ m \in \mathcal{M}_n. \end{aligned}$$

It can be easily seen that problem (10) is concave optimization. Furthermore, since the feasible sets of the constraints are convex sets, there is a zero Lagrange duality gap for problem (10) [25].

2) LAGRANGE DUAL SOLUTION

We relax the UEs QoS constraint C2* and total power constraint C3* by introducing dual variables λ_m and μ_n, respectively, we obtain the Lagrange function as Eq. (11), as shown at the top of next page.

$$L(\mathbf{s},\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu})=\sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{N}_m}\sum_{k\in\mathcal{J}_n}x_{n,m}^k\varepsilon_n B_n\log\left(1+\Gamma\frac{|g_{n,m}^k|^2 s_{n,m}^k}{x_{n,m}^k B_n N_0}\right)-\eta\sum_{n\in\mathcal{N}}\left(\sum_{m\in\mathcal{M}_n}\sum_{k\in\mathcal{J}_n}s_{n,m}^k+\sum_{m\in\mathcal{M}_n}p_{n,m}^c\right)$$
$$-\sum_{m\in\mathcal{M}}\lambda_m\left(R_m^{req}-\sum_{n\in\mathcal{N}_m}\sum_{k\in\mathcal{J}_n}x_{n,m}^k\varepsilon_n B_n\log\left(1+\Gamma\frac{|g_{n,m}^k|^2 s_{n,m}^k}{x_{n,m}^k B_n N_0}\right)\right)-\sum_{n\in\mathcal{N}}\mu_n\left(\sum_{m\in\mathcal{M}_n}\sum_{k\in\mathcal{J}_n}s_{n,m}^k-P_n^{\max}\right). \quad (11)$$

In Eq. (11), $\boldsymbol{\lambda}=[\lambda_m]_{1\times M}$ and $\boldsymbol{\mu}=[\mu_n]_{1\times N}$, then the dual function is given by

$$h(\boldsymbol{\lambda},\boldsymbol{\mu})=\max_{\mathbf{s},\mathbf{x}} L(\mathbf{s},\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu})$$
$$\text{s.t.}\ \sum_{m\in\mathcal{M}_n}x_{n,m}^k\le 1,\ \forall k\in\mathcal{J}_n,\ n\in\mathcal{N},$$
$$0\le x_{n,m}^k\le 1,\ \forall k\in\mathcal{J}_n,\ n\in\mathcal{N},\ m\in\mathcal{M}_n,$$
$$s_{n,m}^k\ge 0,\ \forall n\in\mathcal{N},\ \forall k\in\mathcal{J}_n,\ m\in\mathcal{M}_n, \quad (12)$$

and the dual problem of (10) is

$$\min_{\boldsymbol{\lambda},\boldsymbol{\mu}\ge 0} h(\boldsymbol{\lambda},\boldsymbol{\mu}). \quad (13)$$

Due to the zero Lagrange duality gap between problems (10) and (13), we can solve (12) and (13) for finding the optimum $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$, $\mathbf{s}$ and $\mathbf{x}$.

Based on the Karush-Kuhn-Tucker (KKT) conditions, the following relationship between $\mathbf{s}$ and $\mathbf{x}$ can be obtained

$$s_{n,m}^k=\left[\frac{\varepsilon_n B_n(\lambda_m+\varphi_l+1)}{(\mu_n+\eta)\ln 2}-\frac{B_n N_0}{\Gamma|g_{n,m}^k|^2}\right]^+ x_{n,m}^k, \quad (14)$$

where $[x]^+=\max\{0,x\}$.

Then, substituting Eq. (14) into Eq. (12), extracting the common factor and letting

$$\Lambda_{n,m}^k$$
$$=-\eta\left[\frac{\varepsilon_n B_n(\lambda_m+\varphi_l+1)}{(\mu_n+\eta)\ln 2}-\frac{B_n N_0}{\Gamma|g_{n,m}^k|^2}\right]^+$$
$$+\varepsilon_n B_n\log\left(1+\Gamma\frac{|g_{n,m}^k|^2}{N_0}\left[\frac{\varepsilon_n(\lambda_m+\varphi_l+1)}{(\mu_n+\eta)\ln 2}-\frac{N_0}{\Gamma|g_{n,m}^k|^2}\right]^+\right)$$
$$+\lambda_m\varepsilon_n B_n\log\left(1+\Gamma\frac{|g_{n,m}^k|^2}{N_0}\left[\frac{\varepsilon_n(\lambda_m+\varphi_l+1)}{(\mu_n+\eta)\ln 2}-\frac{N_0}{\Gamma|g_{n,m}^k|^2}\right]^+\right)$$
$$-\mu_n\left[\frac{\varepsilon_n B_n(\lambda_m+\varphi_l+1)}{(\mu_n+\eta)\ln 2}-\frac{B_n N_0}{\Gamma|g_{n,m}^k|^2}\right]^+,$$

we can finally rewrite Eq. (12) as

$$h(\boldsymbol{\lambda},\boldsymbol{\mu})=\max_{\mathbf{x}}\sum_{m\in\mathcal{M}}\sum_{n\in N_m}\sum_{k\in\mathcal{J}_n}\Lambda_{n,m}^k x_{n,m}^k+\mu_n\sum_{n\in\mathcal{N}}P_n^{\max}$$
$$-\eta\sum_{n\in\mathcal{N}}\sum_{m\in\mathcal{M}_n}p_{n,m}^c-\sum_{m\in\mathcal{M}}\lambda_m R_m^{req}, \quad (15)$$

subject to the following constraints:

$$\sum_{m\in\mathcal{M}_n^c}x_{n,m}^k\le 1,\ 0\le x_{n,m}^k\le 1,\ \forall k\in\mathcal{J}_n,\ n\in\mathcal{N}. \quad (16)$$

Eqs. (15) and (16) can be considered as a classical linear assignment problem. The solutions of $x_{n,m}^k$ can either be 0 or 1. Therefore, we found that the optimal solution is binary even after continuity relaxation on $x_{n,m}^k$. More specifically, the allocation of $\mathbf{x}$ is only determined by $\Lambda_{n,m}^k$. Thus for any subcarrier $k\in\mathcal{J}_n, x_{n,m}^k$, where $m=argmax_{m\in M_n}\{\Lambda_{n,m}^k,\forall k\in\mathcal{J}_n, n\in\mathcal{N}\}$, is the best subcarrier allocation.

$$x_{n,m}^k=\begin{cases}1, & m=argmax_{m\in M_n}\{\Lambda_{n,m}^k,\forall k\in\mathcal{J}_n, n\in\mathcal{N}\};\\ 0, & \text{otherwise.}\end{cases} \quad (17)$$

The optimal values of $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ can be solved by using a gradient descent method, i.e.

$$\lambda_m^{\iota+1}=\left[\lambda_m^\iota+\kappa\left(R_m^{req}-\sum_{n\in\mathcal{N}_m}\sum_{k\in\mathcal{J}_n}x_{n,m}^k\varepsilon_n B_n\log(1+\Gamma\frac{|g_{n,m}^k|^2 p_{n,m}^k}{B_n N_0})\right)\right]^+, \quad (18)$$

$$\mu_n^{\iota+1}=\left[\mu_n^\iota+\nu\left(\sum_{m\in\mathcal{M}_n}\sum_{k\in\mathcal{J}_n}x_{n,m}^k p_{n,m}^k-P_n^{\max}\right)\right], \quad (19)$$

where ι stands for the iteration index, κ and ν are sufficiently small positive step-sizes and $[\cdot]^+$ denotes the projection of $[\cdot]$ onto the nonnegative orthant. It is shown in [25] that the gradient descent method is guaranteed to converge to the optimal Lagrange multiplier for some sufficiently small step-size.

A. COMPUTATIONAL COMPLEXITY

At the Inner Loop of **Algorithm 1**, the subgradient method is utilized to solve dual problem (13). In order to achieve δ-optimality, i.e., $|h(\boldsymbol{\lambda},\boldsymbol{\mu})-h(\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)|<\delta$, the number of iterations is on the order of $O(1/\delta^2)$ [26], which does not depend on the number of variables. In each iteration, Eq. (17) needs to be computed for J_n subcarriers. Because that there are $|M_n|$ UEs connected to AP $n\in\mathcal{N}$, Eq. (17) needs to be computed $|M_n|\cdot J_n$ times in each iteration. It can be easily seen that the computational complexity for computing Eq. (17) is $O(|M_n|)$. In resource allocation of AP $n\in\mathcal{N}$, the order of $|\mathcal{M}_n|J_nO(1/\delta^2)$ times is needed to compute Eq. (17), thus the computational complexity for resource allocation of AP $n \in \mathcal{N}$ is $O\left(J_n|\mathcal{M}_n|^2(1/\delta^2)\right)$. Considering that there are N APs in the HetNet, the computational complexity of resource allocation at the Inner Loop is $\sum_{n\in\mathcal{N}} O\left(J_n|\mathcal{M}_n|^2(1/\delta^2)\right)$. At the Outer Loop, the order of $O(1/\epsilon^2)$ iterations is needed for achieving ϵ-optimality, i.e., $|\eta - \eta^{opt}| < \epsilon$. Therefore, the computational complexity of the optimal energy-efficient resource allocation is $\sum_{n\in\mathcal{N}} O\left(J_n|\mathcal{M}_n|^2(1/\delta^2)(1/\epsilon^2)\right)$.

The analysis above shows that the optimal energy-efficient resource allocation has a relatively high computational complexity, it is suitable for the HetNet with high computing capacity. However, if the resource allocation needs to be realized in a short time window, such as 10 ms for LTE system [13], the optimal energy-efficient resource allocation is no longer applicable. For reducing the computational complexity, we will develop a low-complexity algorithm in the next section.

V. LOW COMPLEXITY SUBOPTIMAL RESOURCE ALLOCATION

In this section, we solve the energy-efficient resource allocation problem in two phases: the first is the minimum QoS guaranteed resource allocation (MGRA) and the second is the EE maximization resource allocation (EMRA). Next, we will discuss the two phases in detail.

A. MINIMUM QoS GUARANTEED RESOURCE ALLOCATION

In this phase, we assume that each AP n maximum transmit power $P_n^{\max}$ is divided equally to its subcarriers $k \in \mathcal{J}_n$, and then each UE m is allocated its best subcarrier alternatively until the minimum UEs QoS constraint is guaranteed. In this case, the best subcarrier is defined as the one that maximizes the rate $r_{n,m}^k$. Let $\tilde{\mathcal{J}}_n, \forall n \in \mathcal{N}$ represent the allocated subcarrier set for AP n to grantee the QoS constraint. Let $P_n^{QoS}, \forall n \in \mathcal{N}$ represent the corresponding power used for AP n, i.e. $P_n^{QoS} = \sum_{k\in\tilde{\mathcal{J}}_n} x_{n,m}^k P_n^{\max}/J_n$. Intuitively, the equal power allocation is not the optimal power allocation, and P_n^{QoS} is not the minimum required power for satisfying the QoS constraint. Thus, we further minimize the power allocation for given subcarrier sets $\tilde{\mathcal{J}}_n$, $\forall n \in \mathcal{N}$.

$$\min \sum_{m\in\mathcal{M}} \sum_{n\in\mathcal{N}_m} \sum_{k\in\tilde{\mathcal{J}}_n} p_{n,m}^k \tag{20}$$

subject to the following constraints:

$\overline{\text{C1}}$ (UEs QoS constraint):

$$\sum_{n\in\mathcal{N}_m} \sum_{k\in\tilde{\mathcal{J}}_n} \varepsilon_n B_n \log\left(1 + \Gamma\frac{|g_{n,m}^k|^2 p_{n,m}^k}{B_n N_0}\right) \geq R_m^{req}, \quad \forall m_l \in \mathcal{M}.$$

$\overline{\text{C2}}$ (Total power constraint):

$$\sum_{m\in\mathcal{M}_n} \sum_{k\in\tilde{\mathcal{J}}_n} p_{n,m}^k \leq P_n^{QoS}, \ \forall n \in \mathcal{N},$$

where P_n^{QoS} can be considered as the upper bound of AP n's transmit power.

It is not difficult to verify that problem (20) is convex optimization, which can be solved via convex optimization techniques, such as the interior point method and Lagrange dual method [25].

After solving problem (20), we use R_n^I and P_n^I to represent the current throughput and power consumption of AP n, respectively. The remaining subcarriers are allocated to the connected UEs in a greedy manner, i.e., the subcarrier which has the highest channel gain is allocated to the corresponding UE to maximize the network throughput. The subcarrier set allocated in the greedy manner is represented as $\bar{\mathcal{J}}_n$, $\forall n \in \mathcal{N}$, then $\mathcal{J}_n = \tilde{\mathcal{J}}_n + \bar{\mathcal{J}}_n$.

B. EE MAXIMIZATION RESOURCE ALLOCATION

Intuitively, the constraints $\overline{\text{C1}}$ and $\overline{\text{C2}}$ are already guaranteed in MGRA phase, therefore we do not consider these constraints again in the resource allocation of EMRA phase. For a fixed subcarrier set $\bar{\mathcal{J}}_n$, determined from MGRA phase, we obtain the following EE maximization problem.

$$\max \frac{\sum\limits_{n\in\mathcal{N}} R_n^I + \sum\limits_{n\in\mathcal{N}} \sum\limits_{k\in\bar{\mathcal{J}}_n} \varepsilon_n B_n \log\left(1 + \Gamma\frac{|g_{n,m}^k|^2 p_{n,m}^k}{B_n N_0}\right)}{\sum\limits_{n\in\mathcal{N}} P_n^I + \sum\limits_{m\in\mathcal{M}_n} p_{n,m}^c + \sum\limits_{n\in\mathcal{N}} \sum\limits_{k\in\bar{\mathcal{J}}_n} p_{n,m}^k}$$
$$\text{s.t.} \sum_{k\in\bar{\mathcal{J}}_n} p_{n,m}^k \leq P_n^{\max} - P_n^I, \forall n \in \mathcal{N}, \tag{21}$$

where the constraint shows that the sum of transmit power on the subcarriers in set $\bar{\mathcal{J}}_n$ should be no larger than the remaining maximum allowed transmit power of AP n.

From Eq. (21), we observe that the power allocation of different APs are coupled in the objective function and this coupling will certainly increase the complexity of the EE maximization problem. The following lemma provides a decoupling method which converts optimizing the system EE into maximizing the minimum individual AP EE.

Lemma 3: The optimal EE of problem (21) defined as η_{EE}^{II} is lower bounded by

$$\min_{n\in\mathcal{N}} \left\{ \max_{p_{n,m}^k, k\in\bar{\mathcal{J}}_n} \frac{R_n^I + \sum\limits_{k\in\bar{\mathcal{J}}_n} \varepsilon_n B_n \log\left(1 + \Gamma\frac{|g_{n,m}^k|^2 p_{n,m}^k}{B_n N_0}\right)}{P_n^I + \sum\limits_{m\in\mathcal{M}_n} p_{n,m}^c + \sum\limits_{k\in\bar{\mathcal{J}}_n} p_{n,m}^k} \right\}.$$

Proof: Similar proof can be found in [27]. ■

The lemma above enables us to split the joint and complex optimization (21) into maximizing the individual EE of AP n.

$$\max \frac{R_n^I + \sum_{k\in\bar{\mathcal{J}}_n} \varepsilon_n B_n \log\left(1 + \Gamma\frac{|g_{n,m}^k|^2 p_{n,m}^k}{B_n N_0}\right)}{P_n^I + \sum_{m\in\mathcal{M}_n} p_{n,m}^c + \sum_{k\in\bar{\mathcal{J}}_n} p_{n,m}^k}$$
$$\text{s.t.} \sum_{k\in\bar{\mathcal{J}}_n} p_{n,m}^k \leq P_n^{\max} - P_n^I. \tag{22}$$

Intuitively, optimization problem (22) is a fractional programming. However, using the transformation method in Section IV may has a relatively high computational complexity. In this subsection, we propose a two-lever algorithm. At the inner loop, we solve a throughput maximization problem of AP n at the given total power P_n, where $0 \le P_n \le P_n^{\max} - P_n^I$. At the outer loop, we solve the EE maximization problem of power P_n. Next, we will discuss the two loops in detail.

1) THROUGHPUT MAXIMIZATION

Given the total transmit power P_n, maximizing the individual EE is equivalent to maximizing the AP throughput, i.e.,

$$\max \sum_{k\in\bar{\mathcal{J}}_n} \varepsilon_n B_n \log\left(1+\Gamma\frac{|g_{n,m}^k|^2 p_{n,m}^k}{B_n N_0}\right)$$
$$\text{s.t} \sum_{k\in\bar{\mathcal{J}}_n} p_{n,m}^k \le P_n. \tag{23}$$

Problem (23) is a concave optimization problem, and many iteration algorithms, such as interior point method and Lagrange dual method [25], can be used to solve it. However, the iteration algorithm will increase the computational complexity of the throughput maximization problem. Therefore, we will propose a low-complexity resource allocation method in this subsection.

The Lagrange function of problem (23) is given as

$$L(\mathbf{p},\beta) = \sum_{k\in\bar{\mathcal{J}}_n} \varepsilon_n B_n \log\left(1+\Gamma\frac{|g_{n,m}^k|^2 p_{n,m}^k}{B_n N_0}\right) - \beta\left(\sum_{k\in\bar{\mathcal{J}}_n} p_{n,m}^k - P_n\right). \tag{24}$$

By computing the derivation $\frac{\partial L}{\partial p_{n,m}^k} = 0$, we obtain the following equation

$$p_{n,m}^k + \frac{B_n N_0}{\Gamma g_{n,m}^k} = \frac{\varepsilon_n B_n}{\beta \ln 2}. \tag{25}$$

The right-hand-side of Eq. (25) is not related with the subcarrier indicator k, then, for any two subcarriers $k_1, k_2 \in \bar{\mathcal{J}}_n$, the following relationship can be derived

$$p_{n,m}^{k_2} = p_{n,m}^{k_1} + \frac{B_n N_0}{\Gamma}\left(\frac{1}{g_{n,m}^{k_1}} - \frac{1}{g_{n,m}^{k_2}}\right). \tag{26}$$

Eq. (26) shows that as long as the transmit power on a certain subcarrier is determined, the transmit powers on other subcarriers can be computed. More specifically, the sum of the transmit powers on these subcarriers is

$$\sum_{k\in\bar{\mathcal{J}}_n} p_{n,m}^k = \Theta_n p_{n,m}^{k_1} + \Theta_n \frac{B_n N_0}{g_{n,m}^{k_1}\Gamma} - \sum_{k\in\bar{\mathcal{J}}_n} \frac{B_n N_0}{g_{n,m}^k \Gamma}, \tag{27}$$

where $\Theta_n = |\bar{\mathcal{J}}_n|$.

The constraint in problem (23) shows that $\sum_{k\in\bar{\mathcal{J}}_n} p_{n,m}^k \le P_n$, then $p_{n,m}^{k_1}$ can be calculated as

$$p_{n,m}^{k_1} \le \frac{1}{\Theta_n}\left(P_n + \sum_{k\in\bar{\mathcal{J}}_n} \frac{B_n N_0}{g_{n,m}^k \Gamma} - \frac{\Theta_n B_n N_0}{g_{n,m}^{k_1}\Gamma}\right). \tag{28}$$

Taking the upper bound of Eq. (28), we can compute the transmit power on subcarrier k_1. Then, by using Eq. (26), we can compute the transmit powers of all the remaining subcarriers in set $\bar{\mathcal{J}}_n$. However, the computed transmit power may be illegal value, such as $p_{n,m}^{k_1} < 0$. For avoiding this case, we firstly arrange the subcarriers as $g_{n,m}^{k_1} \le g_{n,m}^{k_2} \le \cdots \le g_{n,m}^{k_{\Theta_n}}$, then $p_{n,m}^{k_1} \le p_{n,m}^{k_2} \le \cdots \le p_{n,m}^{k_{\Theta_n}}$. Taking the upper bound of Eq. (28), we compute $p_{n,m}^{k_1}$ as

$$p_{n,m}^{k_1} = \frac{1}{\Theta_n}\left(P_n + \sum_{k\in\bar{\mathcal{J}}_n} \frac{B_n N_0}{g_{n,m}^k \Gamma} - \frac{\Theta_n B_n N_0}{g_{n,m}^{k_1}\Gamma}\right). \tag{29}$$

If $p_{n,m}^{k_1} < 0$, we set the transmit power of this subcarrier as 0, remove it from $\bar{\mathcal{J}}_n$ and compute the transmit power on $g_{n,m}^{k_2}$

$$p_{n,m}^{k_2} = \frac{1}{\Theta_n}\left(P_n + \sum_{k\in\bar{\mathcal{J}}_n/k_1} \frac{B_n N_0}{g_{n,m}^k \Gamma} - \frac{(\Theta_n - 1) B_n N_0}{g_{n,m}^{k_2}\Gamma}\right), \tag{30}$$

until find that $p_{n,m}^k \ge 0$. Afterwards, the transmit powers on other subcarriers can be computed by Eq. (26).

Therefore, the throughput maximization problem can be solved, and the optimal value can be considered as function of P_n, we denote it by $R_n^{\max}(P_n)$.

2) EE MAXIMIZATION

We define the individual EE of AP n with respect to the given total transmit power P_n as $\eta_n(P_n) = \frac{R_n^I + R_n^{\max}(P_n)}{P_n^I + P_n + \sum_{m_l\in\mathcal{M}_n} p_{n,m_l}^c}$. Then, EE maximization problem (21) is reformulated as

$$\eta_n^{\max} = \max_{0\le P_n \le P_n^{\max} - P_n^I} \eta_n(P_n). \tag{31}$$

The following lemma shows that $\eta_n(P_n)$ is a quasiconcavity function with respect to P_n.

Lemma 4: $\eta_n(P_n)$ is a quasiconcavity function in total transmit power P_n.

Proof: Please refer to Appendix A for the proof. ■

For any quasiconcavity function, there is always a unique globally optimal solution [27]. Thus, we give the following theorem to show the properties of the globally optimal solution.

Theorem 1: If $\eta_n(P_n)$ is a quasiconcavity function, there exists a unique globally optimal solution for problem (31), and $\eta_n(P_n)$ is maximized at

1) $P_n = 0$, *if* $\left.\frac{d\eta_n(P_n)}{dP_n}\right|_{P_n=0} \le 0$;
2) $P_n = P_n^*$, *where* $\left.\frac{d\eta_n(P_n)}{dP_n}\right|_{P_n=P_n^*} = 0$, *if* $\left.\frac{d\eta_n(P_n)}{dP_n}\right|_{P_n=0} \ge 0$ *and* $\left.\frac{d\eta_n(P_n)}{dP_n}\right|_{P_n=P_n^{\max}-P_n^I} \le 0$;

3) $P_n = P_n^{\max} - P_n^I$, *if* $\left.\frac{d\eta_n(P_n)}{dP_n}\right|_{P_n = P_n^{\max} - P_n^I} \geq 0$,

where $\frac{d\eta_n(P_n)}{dP_n} = \frac{R_n^{\max}(P_n)'(P_n^I + P_n + \sum_{m_l \in \mathcal{M}_n} p_{n,m_l}^c) - (R_n^I + R_n^{\max}(P_n))}{(P_n^I + P_n + \sum_{m_l \in \mathcal{M}_n} p_{n,m_l}^c)^2}$

and $R_n^{\max}(P_n)' = \frac{dR_n^{\max}(P_n)}{dP_n}$.

According to **Theorem 1**, the EE maximization problem (31) can be solved by a bisection algorithm. However, it is so difficult to obtain $R_n^{\max}(P_n)'$ due to the fact that $R_n^{\max}(P_n)$ can not be expressed as an explicit function of P_n. For solving this difficult, we give the following lemma.

Lemma 5: For any $P_n > 0$, the derivative of $R_n^{\max}(P_n)$ is

$$R_n^{\max}(P_n)' = \max_{k \in \bar{\mathcal{J}}_n} \frac{\varepsilon_n B_n \Gamma |g_{n,m_l}^k|^2}{(B_n N_0 + \Gamma |g_{n,m_l}^k|^2 p_{n,m_l}^k) \ln 2}.$$

Proof: Please refer to Appendix B for the proof. ■

Then, the EE maximization problem (31) can be effectively solved by a bisection algorithm.

Finally, the framework of low-complexity suboptimal resource allocation algorithm is given in **Algorithm 2**.

Algorithm 2 Low-Complexity Suboptimal Resource Allocation

MGRA:

1) Divide APs' maximum transmit power equally to its subcarriers, and then allocate the best subcarrier alternatively to the corresponding UEs until both the minimum UEs QoS constraint and SLA contract constraint are guaranteed.
2) For given subcarrier set $\tilde{\mathcal{J}}_n$, $\forall n \in \mathcal{N}$, solve optimization problem (20).
3) Greedily allocate APs' remaining subcarriers to the connected UEs.

EMRA:

1) For given subcarrier set $\bar{\mathcal{J}}_n$ and total transmit power P_n, solve throughput maximization problem (23) (Inner loop).
2) Solve EE maximization problem (31) with the bisection-based method given in **Algorithm 3** (Outer loop).

Algorithm 3 Bisection-Based Method

1) Initialize the lower and upper bound of transmit power, $P_n^{low} = 0$ and $P_n^{up} = P_n^{\max} - P_n^I$.
2) Calculate $d = \frac{d\eta_n(P_n)}{dP_n}$, where $P_n = (P_n^{low} + P_n^{up})/2$.
3) If $d > 0$, then update $P_n^{low} = P_n$. If $d < 0$, then update $P_n^{up} = P_n$.

C. COMPUTATIONAL COMPLEXITY

At the MGRA phase, the computational complexity is mainly related to the iteration algorithm for solving problem (20). If the δ-optimality wants to be achieved, the computational complexity of the subgradient algorithm is on the order of $O(1/\delta^2)$. At the outer loop of EMRA phase, bisection method is used to obtain the optimal transmit power, whose computational complexity is relatively low. If the required accuracy is υV, the order of iteration is $O(\log_2(1/\upsilon))$, where V is the difference between upper bound and lower bound of the initialized EE. At the inner loop of EMRA phase, the computational complexity comes from computing the transmit powers in Eqs. (29), (30) and (26), whose complexity is $O(\Theta_n)$. Thus, the computational complexity of EMRA phase is $O(\Theta_n \log_2(1/\upsilon))$. Together with MGRA phase, computational complexity of suboptimal resource allocation is $O(1/\delta^2)+\sum_{n\in\mathcal{N}} O(\Theta_n \log_2(1/\upsilon))$, which is much lower than the optimal resource allocation.

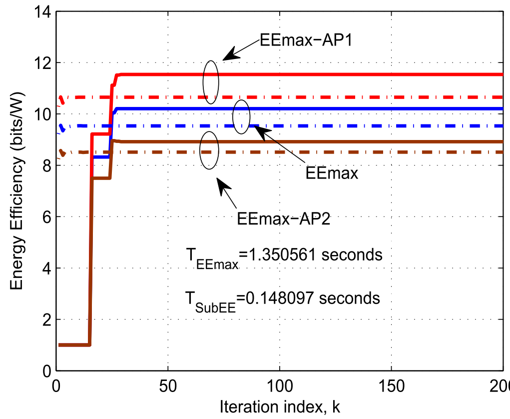


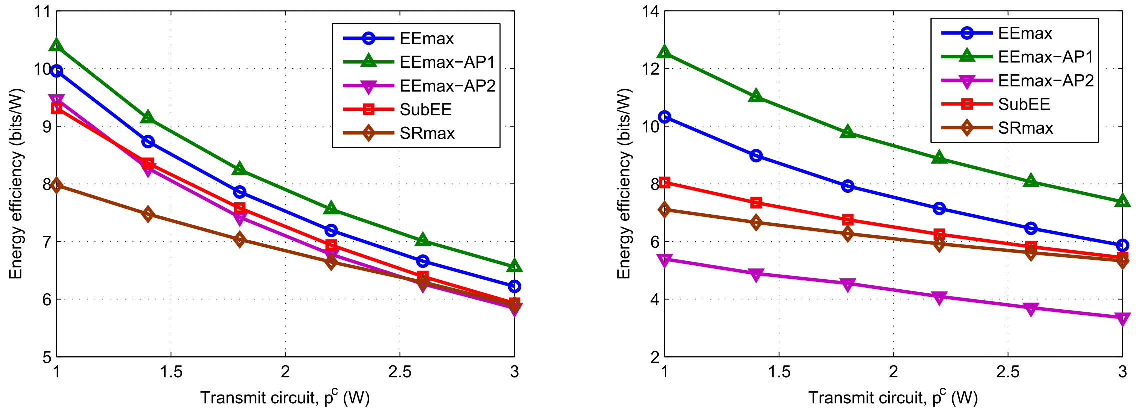
FIGURE 2. Convergence results of EE under EEmax and SubEE algorithms. Solid and dotted lines represent EEmax and SubEE algorithms respectively. The minimum rate requirements of the four UEs are 20, 30, 25 and 18bits/s, respectively. The network efficiencies $\varepsilon_1 = 0.8$ and $\varepsilon_2 = 0.8$. The circuit power parameters are taken as $p_{n,m}^c = 2$, $\forall n \in \mathcal{N}$, $m_l \in \mathcal{M}$. The maximum transmit power of APs are $P_1^{\max} = P_2^{\max} = 10$W.

VI. SIMULATION RESULTS

In this section, we present experiments to evaluate the performance of the optimal energy-efficient resource allocation algorithm (EEmax) and the low complexity suboptimal resource allocation algorithm (SubEE). In particular, we compare our proposed designs to the throughput maximization strategy (SRmax), which is widely used for the resource allocation in wireless networks [10], [13], [14]. We note that the SRmax problem is a convex optimization, which can be solved efficiently by the standard subgradient method.

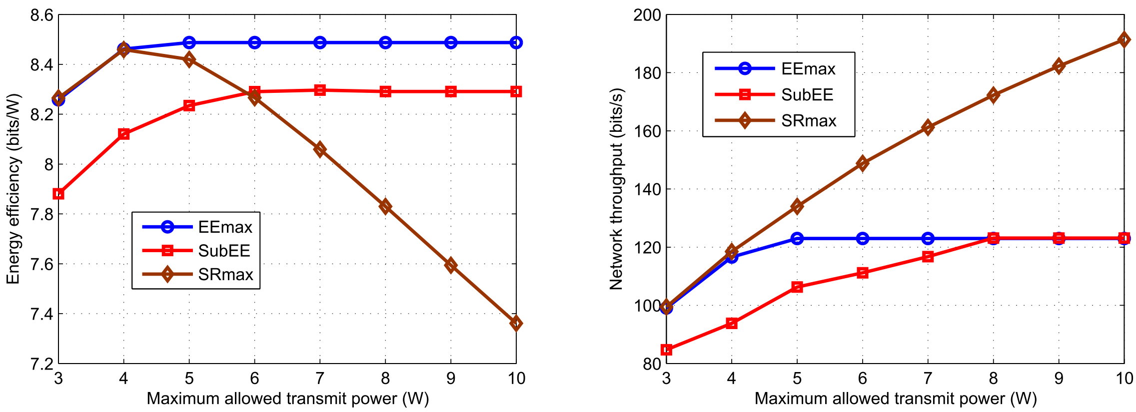
During the simulation, we use the HetNet having two APs and four UEs in the network topology. For simplicity, we assume that both the subcarrier spacing B_n and the power spectral density of additive white Gaussian noise N_0 are normalized to unit [28]. We model the channel gain as Gaussian random variables in [7] and [24]. The Shannon capacity gap is 0.7. The other values will be specified in each numerical experiment.

In the first experiment, we investigate the convergence results of EEmax and SubEE algorithms in Fig. 2. In the figure, T_{EEmax} and T_{SubEE} represent the elapsed time of



(a) (b)

FIGURE 3. EE versus circuit power. The minimum rate requirements of the four UEs are 20, 30, 25 and 18bits/s, respectively. The maximum transmit power of APs are $P_1^{max} = P_2^{max} = 10$W. (a) Network efficiencies $\varepsilon_1 = 0.8$ and $\varepsilon_2 = 0.8$. (b) Network efficiencies $\varepsilon_1 = 0.9$ and $\varepsilon_2 = 0.5$.



(a) (b)

FIGURE 4. EE and throughput versus maximum allowed transmit power. The minimum rate requirements of the four UEs are 20, 18, 22 and 18bits/s, respectively. The network efficiencies are $\varepsilon_1 = 0.8$ and $\varepsilon_2 = 0.8$. (a) EE versus maximum allowed transmit power. (b) Throughput maximum allowed transmit power.

EEmax and SubEE algorithms respectively. The codes are executed on a 64-bit Windows 7 operating system with AMD Athlon II x4631 Quad-Core Processor, 4 Gbyte RAM. Obviously, SubEE algorithm is more computationally efficient than EEmax algorithm. However, SubEE algorithm achieves smaller system EE and individual AP EE than EEmax algorithm, but with much lower complexity.

We test the EE versus circuit power under various resource allocation algorithms in Fig. 3. The figure also plots the EE of individual AP to investigate the effect of network efficiency. We remark that the system EE of HetNet defined in this paper is not equal to the sum of the EE of individual APs. We observe that the difference in terms of EE of the two APs is remarkable in Fig. 3 (b), which is the result of a large difference in network efficiencies. The figure shows that SubEE algorithm achieves smaller EE than EEmax algorithm but larger than SRmax algorithm. We also observe that EEmax and SubEE algorithms achieve remarkably higher EE than SRmax algorithm when the circuit power is small. As the circuit power increases, the EE of all algorithms in comparison is reduced. This can be explained as follows. If the circuit power is small, the main contributing term of the denominator in Eq. (6) is the transmit power. As the rate function is logarithmic function corresponding to transmit power, the EEmax and SubEE algorithms utilize a small transmit power to obtain the achievable EE. On the other hand, if the circuit power is large, the main contributing term of the denominator in Eq. (6) is the circuit power. Then, maximizing EE is equivalent to maximizing the sum rate. That is the why the EE of all algorithms in comparison is reduced when the circuit power is large.

To further evaluate the impact of maximum allowed transmit power of APs on EE and throughput performance, Fig. 4 compares the EE and throughput performance of different algorithms in terms of $P^{\max}$. In this experiment, we set $P_1^{\max} = P_2^{\max} = P^{\max}$. When $P^{\max}$ is small, the

EE performance increases with the rising of $P^{\max}$ for all algorithms. Fig. 4 (a) even shows that SRmax algorithm achieves the EE performance similar to EEmax and larger than SubEE when $P^{\max}$ is small. When $P_n^{\max}$ is large, the EE performance under EEmax and SubEE algorithms keeps stable, while the one under SRmax algorithm drops rapidly. That is because both EEmax and SubEE algorithms find the optimal maximum allowed transmit power for APs, which is probably around 5-6w for EEmax algorithm and 7-8W for SubEE algorithm, respectively. Therefore, the consumed transmit powers under EEmax and SubEE algorithms are no longer growth, even through the maximum allowed transmit power increases. However, for maximizing the system throughput, SRmax algorithm greedily consumes all of the allowed transmit power, hence it incurs terrible EE performance. Fig. 4 (b) shows that SRmax algorithm always achieves the maximum throughput among all of the algorithms, and it increases almost linearly with the rising of $P^{\max}$. However, the throughput under EEmax and SubEE algorithms keeps stable when the maximum allowed transmit power is larger than the corresponding optimal value. Therefore, the EE performance under EEmax and SubEE algorithms keeps stable when $P_n^{\max}$ is large.

VII. CONCLUSIONS

In this paper, we investigated the energy-efficient resource allocation problem for the HetNet with multi-homed UEs. We formulated the resource optimization as a non-concave EE maximization problem, which was converted by a fractional programming theory into an MINO problem. Then, we solved the MINO problem by using the continuity relaxation and Lagrange dual method. Afterwards, we developed an optimal energy-efficient joint power and subcarrier allocation algorithm. For reducing the computational complexity of the resource allocation problem, we also proposed a two-phase optimization method, which included the minimum rate guaranteed resource allocation and the energy-efficient resource allocation. We showed that the low-complexity algorithm achieved the suboptimal EE with significantly lower computational complexity compared with the optimal one.

APPENDIX A
PROOF OF LEMMA 4

Proof: Denote the superlevel set of $\eta_n(P_n)$ as $S_\alpha = \{P_n | \eta_n(P_n) \geq \alpha, P_n \geq 0\}$, where α is a positive value because of the nonnegative characteristic of EE. The superlevel set S_α can be also denoted as

$$\{P_n | R_n^I + R_n^{\max}(P_n) - \alpha P_n^I - \alpha P_n - \alpha \sum_{m_l \in \mathcal{M}_n} p_{n,m_l}^c \geq 0, P_n \geq 0\}$$

Intuitively, $R_n^{\max}(P_n)$ is a strictly concave function in P_n, thus there exist $P_n^1 \in S_\alpha$ and $P_n^2 \in S_\alpha$ satisfying

$$R_n^{\max}(\tau P_n^1 + (1-\tau)P_n^2) \geq \tau R_n^{\max}(P_n^1) + (1-\tau) R_n^{\max}(P_n^2).$$

There comes that

$$\begin{aligned}
& R_n^I + R_n^{\max}(\tau P_n^1 + (1-\tau)P_n^2) - \alpha P_n^I - \alpha(\tau P_n^1 + (1-\tau)P_n^2) \\
& \quad - \alpha \sum_{m_l \in \mathcal{M}_n} p_{n,m_l}^c \geq R_n^I + \tau R_n^{\max}(P_n^1) + (1-\tau) R_n^{\max}(P_n^2) \\
& \quad - \alpha P_n^I - \alpha(\tau P_n^1 + (1-\tau)P_n^2) - \alpha \sum_{m_l \in \mathcal{M}_n} p_{n,m_l}^c \\
& = (1-\tau)(R_n^I + R_n^{\max}(P_n) - \alpha P_n^I - \alpha P_n - \alpha \sum_{m_l \in \mathcal{M}_n} p_{n,m_l}^c) \\
& \quad + \tau(R_n^I + R_n^{\max}(P_n) - \alpha P_n^I - \alpha P_n - \alpha \sum_{m_l \in \mathcal{M}_n} p_{n,m_l}^c) = 0.
\end{aligned}$$

Thus, $\tau P_n^1 + (1-\tau)P_n^2 \in S_\alpha$, and S_α is a convex set. Hence, $\eta_n(P_n)$ is a quasiconcavity function in total transmit power P_n. ■

APPENDIX B
PROOF OF LEMMA 5

Proof: According to the definition of $R_n^{\max}(P_n)$, we obtain that

$$\begin{aligned}
& R_n^{\max}(P_n + \Delta P) - R_n^{\max}(P_n) \\
& \quad = \max_{k \in \bar{\mathcal{J}}_n} \varepsilon_n B_n \log\left(1 + \Gamma \frac{|g_{n,m_l}^k|^2 (p_{n,m_l}^k + \Delta P)}{B_n N_0}\right) \\
& \qquad - \varepsilon_n B_n \log\left(1 + \Gamma \frac{|g_{n,m_l}^k|^2 p_{n,m_l}^k}{B_n N_0}\right).
\end{aligned}$$

Using the equation of

$$R_n^{\max}(P_n)' = \lim_{\Delta P \to 0} \frac{R_n^{\max}(P_n + \Delta P) - R_n^{\max}(P_n)}{\Delta P},$$

we obtain

$$R_n^{\max}(P_n)' = \max_{k \in \bar{\mathcal{J}}_n} \frac{\varepsilon_n B_n \Gamma |g_{n,m_l}^k|^2}{(B_n N_0 + \Gamma |g_{n,m_l}^k|^2 p_{n,m_l}^k) \ln 2}.$$

■

REFERENCES

[1] I. Hwang, B. Song, and S. S. Soliman, "A holistic view on hyper-dense heterogeneous and small cell networks," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 20–27, Jun. 2013.
[2] C. Xu, T. Liu, J. Guan, H. Zhang, and G.-M. Muntean, "CMT-QA: Quality-aware adaptive concurrent multipath data transfer in heterogeneous wireless networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 11, pp. 2193–2205, Nov. 2013.
[3] J. Wu, B. Cheng, C. Yuen, Y. Shang, and J. Chen, "Distortion-aware concurrent multipath transfer for mobile video streaming in heterogeneous wireless networks," *IEEE Trans. Mobile Comput.*, vol. 14, no. 4, pp. 688–701, Apr. 2015.
[4] R. Kuntz, J. Montavont, and T. Noel, "Multihoming in IPv6 mobile networks: Progress, challenges, and solutions," *IEEE Commun. Mag.*, vol. 51, no. 1, pp. 128–135, Jan. 2013.
[5] J. Jiang, M. Peng, K. Zhang, and L. Li, "Energy-efficient resource allocation in heterogeneous network with cross-tier interference constraint," in *Proc. IEEE PIMRC Workshops*, Sep. 2013, pp. 168–172.
[6] J. Wu, Y. Zhang, M. Zukerman, and E. K. N. Yung, "Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 803–826, 2nd Quart., 2015.
[7] H. Ju, B. Liang, J. Li, Y. Long, and X. Yang, "Adaptive cross-network cross-layer design in heterogeneous wireless networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 655–669, Feb. 2015.
[8] K. Nakauchi, Y. Shoji, and N. Nishinaga, "Airtime-based resource control in wireless LANs for wireless network virtualization," in *Proc. 4th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Jul. 2012, pp. 166–169.

[9] Q. Zhu and X. Zhang, "Bayesian-game based power and spectrum virtualization for maximizing spectrum efficiency over mobile cloud-computing wireless networks," in *Proc. IEEE INFOCOM WKSHPS*, Apr./May 2015, pp. 378–383.

[10] H. Zhang, W. Wang, X. Li, and H. Ji, "User association scheme in cloud-ran based small cell network with wireless virtualization," in *Proc. INFOCOM WKSHPS*, Apr. 2015, pp. 384–389.

[11] J. Qiu *et al.*, "Hierarchical resource allocation framework for hyper-dense small cell networks," *IEEE Access*, vol. 4, pp. 8657–8669, 2016.

[12] G. Liu, F. R. Yu, H. Ji, and V. C. M. Leung, "Virtual resource management in green cellular networks with shared full-duplex relaying and wireless virtualization: A game-based approach," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 7529–7542, Sep. 2016.

[13] M. I. Kamel, L. B. Le, and A. Girard, "LTE multi-cell dynamic resource allocation for wireless network virtualization," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, New Orleans, LA, USA, Mar. 2015, pp. 966–971.

[14] L. Chen, F. R. Yu, H. Ji, G. Liu, and V. C. M. Leung, "Distributed virtual resource allocation in small-cell networks with full-duplex self-backhauls and virtualization," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5410–5423, Jul. 2016.

[15] Q. Yu, G. Wu, Q. Tang, and S. Li, "Low-complexity energy-efficient resource allocation in ofdma networks," in *Proc. WCSP*, Oct. 2012, pp. 1–5.

[16] M. Ismail, W. Zhuang, and S. Elhedhli, "Energy and content aware multi-homing video transmission in heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3600–3610, Jul. 2013.

[17] N. Abbas, Z. Dawy, H. Hajj, and S. Sharafeddine, "Energy-throughput tradeoffs in cellular/wifi heterogeneous networks with traffic splitting," in *Proc. IEEE WCNC*, Apr. 2014, pp. 2294–2299.

[18] M. Ismail and W. Zhuang, "Mobile terminal energy management for sustainable multi-homing video transmission," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4616–4627, Aug. 2014.

[19] C. Yang, J. Yue, M. Sheng, and J. Li, "Tradeoff between energy-efficiency and spectral-efficiency by cooperative rate splitting," *J. Commun. Netw.*, vol. 16, no. 2, pp. 121–129, Apr. 2014.

[20] Q.-D. Vu, L.-N. Tran, M. Juntti, and E.-K. Hong, "Energy-efficient bandwidth and power allocation for multi-homing networks," *IEEE Trans. Signal Process.*, vol. 63, no. 7, pp. 1684–1699, Apr. 2015.

[21] C. He, B. Sheng, P. Zhu, X. You, and G. Y. Li, "Energy- and spectral-efficiency tradeoff for distributed antenna systems with proportional fairness," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 894–902, May 2013.

[22] C. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1982.

[23] J.-P. Crouzeix and J. A. Ferland, "Algorithms for generalized fractional programming," *Math. Program.*, vol. 52, no. 1, pp. 191–207, May 1991.

[24] W. Wu, Q. Yang, P. Gong, and K. S. Kwak, "Energy-efficient resource optimization for OFDMA-based multi-homing heterogenous wireless networks," *IEEE Trans. Signal Process.*, vol. 64, no. 22, pp. 5901–5913, Nov. 2016.

[25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.:Cambridge Univ. Press, Mar. 2004.

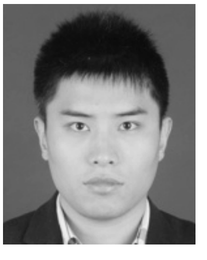
[26] N. Mokari, M. R. Javan, and K. Navaie, "Cross-layer resource allocation in OFDMA systems for heterogeneous traffic with imperfect CSI," *IEEE Trans. Veh. Technol.*, vol. 59, no. 2, pp. 1011–1017, Feb. 2010.

[27] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient resource allocation in OFDMA networks," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3767–3778, Dec. 2012.

[28] L. Song, Z. Han, Z. Zhang, and B. Jiao, "Non-cooperative feedback-rate control game for channel state information in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 1, pp. 188–197, Jan. 2012.

RUNZI LIU (M'16) received the B.S. degree (Hons.) in telecommunications engineering and the Ph.D. degree in communication and information systems from Xidian University, Xi'an, China, in 2011 and 2016, respectively. Since 2016, she has been with the School of Telecommunications Engineering, Xidian University, where she is currently an Assistant Professor. Her research interests include resource management and performance analysis in wireless networks and space networks.

MIN SHENG (M'03–SM'16) received the M.S. and Ph.D. degrees in communication and information systems from Xidian University in 2000 and 2004, respectively. She is currently a Full Professor with the State Key Laboratory of Integrated Service Networks, Xidian University. Her general research interests include mobile ad hoc networks, 5G mobile communication systems, and satellite communications networks. She was named as a Changjiang Scholar by the Ministry of Education, China.



WEIHUA WU received the B.S. and M.E. degrees in telecommunications engineering and the Ph.D. degree in communication and information systems from Xidian University, China, in 2011, 2014, and 2017, respectively. Since 2018, he has been with the School of Telecommunications Engineering, Xidian University, where he is currently an Assistant Professor. His research interests include wireless resource allocation and stochastic network optimization and their applications in multihoming wireless networks.